# Crystal structure of a bacterial chitinase at 2.3 Å resolution

**Anastassis Perrakis[1]*, Ivo Tews[1], Zbigniew Dauter[1], Amos B Oppenheim[2], Ilan Chet[3], Keith S Wilson[1] and Constantin E Vorgias[1]***

[1]European Molecular Biology Laboratory, c/o DESY Notkestrasse 85, 22603 Hamburg, Germany, [2]Department of Molecular Genetics, Hebrew University-Hadassah Medical School, Jerusalem, Israel and [3]The Otto Warburg Center for Biotechnology, Faculty of Agriculture, Hebrew University, Rehovot, Israel

**Background:** Chitinases cleave the β-1-4-glycosidic bond between the *N*-acetyl-D-glucosamine units of which chitin is comprised. Chitinases are present in plants, bacteria and fungi, but whereas structures are available for two prototypic plant enzymes, no structure is available for a bacterial or fungal chitinase.

**Results:** To redress this imbalance, the structure of native chitinase A from *Serratia marcescens* has been solved by multiple isomorphous replacement and refined at 2.3 Å resolution, resulting in a crystallographic R-factor of 16.2%. The enzyme comprises three domains: an all-β-strand amino-terminal domain, a catalytic α/β-barrel domain, and a small α+β-fold domain. There are several residues with unusual geometries in the structure. Structure determination of chitinase A in complex with $N,N',N'',N'''$-tetra-acetylo-chitotetraose, together with biochemical and sequence analysis data, enabled the positions of the active-site and catalytic residues to be proposed.

**Conclusions:** The reaction mechanism seems to be similar to that of lysozyme and most other glycosylhydrolases, i.e. general acid-base catalysis. The role of the amino-terminal domain could not be identified, but it has similarities to the fibronectin III domain. This domain may possibly facilitate the interaction of chitinase A with chitin.

**Structure** 15 December 1994, **2**:1169–1180

Key words: α/β-barrel, chitin, crystallography, structure

## Introduction

Chitinases (EC 3.2.1.14) hydrolyze the chitin homopolymer into smaller oligomeric saccharides. Chitin, one of the most abundant natural polymers, comprises multiple *N*-acetyl-D-glucosamine (NAG) units. It is the main structural component of the cell wall of most fungi and of the exoskeleton of arthropods.

Chitinases are widespread in nature and have been found in bacteria, fungi, plants, invertebrates (mainly nematodes, insects and crustaceans) and all classes of vertebrates. The roles of chitinases in these organisms are diverse and subject to intensive investigation. In vertebrates, chitinases are usually part of the digestive tract [1]. In insects and crustaceans, chitinases are associated with the need for partial degradation of old cuticle, and their secretion is controlled by a complex hormonal mechanism [2]. In plants, a major part of the inducible defence response against exogenous pathogenesis is the secretion of proteinase inhibitors, glucanases and chitinases [3–5]. Chitinases in fungi are thought to have autolytic, nutritional and morphogenetic roles [6,7]. Finally, bacteria secrete chitinases that catalyze the degradation of chitin, the products of which act as a source of energy.

Chitinases constitute families 18 and 19 of the glycosylhydrolases [8,9]. Family 19 is in general highly conserved in primary structure and contains only plant chitinases. Family 18 is diverse in evolutionary terms [10] and contains enzymes from plants, bacteria and fungi. The structure of a family 19 enzyme, from *Hordeum vulgare*, has been solved at 2.8 Å resolution [11]. The structure of a family 18 enzyme, hevamine, with combined chitinase and lysozyme activities, from the plant *Hevea brasiliensis* has also been reported [12]. No structure of a bacterial or fungal chitinase is available.

The biochemical properties of chitinase A (ChiA), a family 18 enzyme from the Gram-negative soil bacterium *Serratia marcescens*, have been previously described [13]. The enzyme has optimum activity at pH 4.5–7.0. The main product of chitin cleavage is diacetylchitobiose and, for higher enzyme/substrate ratios, free NAG. The enzyme acts as an exochitinase and excises oligosaccharides from the non-reducing end of the polysaccharide chain. A Michaelis constant ($K_m$) of 4.4 mM (expressed in NAG equivalents), was reported for the enzyme, which was shown to be highly specific for chitin, above all other yeast cell-wall components.

The gene encoding chitinase A was previously isolated, cloned and sequenced [14,15]. A plasmid carrying the ChiA gene under the control of the oLpL operator and promoter was introduced into *Escherichia coli* to yield the ChiA-overproducing strain A5745, used throughout this study [16]. The enzyme is secreted into the culture medium. An efficient purification protocol has been previously established [17]. The purified enzyme is processed, as confirmed by microsequencing of the amino-terminal end, and the leader peptide, which is 23 amino acids long, is cleaved (the residue numbering used

*Corresponding authors.

 ISSN 0969-2126

throughout this study starts with residue 24). The mature enzyme comprises 540 amino acids, and has a calculated molecular weight of 58.7 kDa. On SDS-PAGE it appears as a single band of 58–60 kDa.

Microbial chitinases are important agents in resistance against the fungal infection of plants, including some infections of major economic interest. ChiA is the most studied bacterial chitinase on account of its applications in biological control. Recent studies showed that, firstly, the infection of bean and cotton plants by the soil fungi *Rhizoctonia solani* and *Sclerotium rolfsii* could be effectively reduced by the presence of the *E. coli* ChiA-overproducing strain [18], and, secondly, expression of the ChiA gene in transgenic tobacco plants improved resistance of the plants to fungal infection [19].

## Results and discussion

### Electron-density map and model building

The structure of ChiA was solved by multiple isomorphous replacement (MIR). The MIR map was solvent flattened and the electron density was of good quality (Fig. 1). Refinement of the initial model and addition of 332 solvent molecules resulted in a final R-factor of 16.2% for a model with very good geometry.

### Primary structure

In the SWISSPROT database [20], a file reporting two independently determined sequences was found (accession no. P07254). The older sequence [14] had 11 differences (including an insertion of one amino acid) compared with the most recent one (JC Koo *et al.*, unpublished data). During model building, the electron density fitted excellently to the second sequence.

While building the last region of the model, a sequence of 20 residues, residues 410–429, was identified for which the side chains did not fit the density. The main-chain density was very well defined in that region, and it was clear that 21 residues rather that 20 should be fitted in that density. It was decided to sequence this area in order to resolve possible GC compressions. This part of the gene was copied by polymerase chain reaction (PCR), cloned into the pcrII™ vector (Invitrogen) and sequenced. The resulting sequence is in perfect agreement with the electron density (Fig. 2).

The final electron-density map shows clearly that residues Ala73 and Ala351 are more likely to be valine or threonine. In addition, residue 395 is proline, not alanine. These differences are probably attributable to mistakes in DNA sequencing.

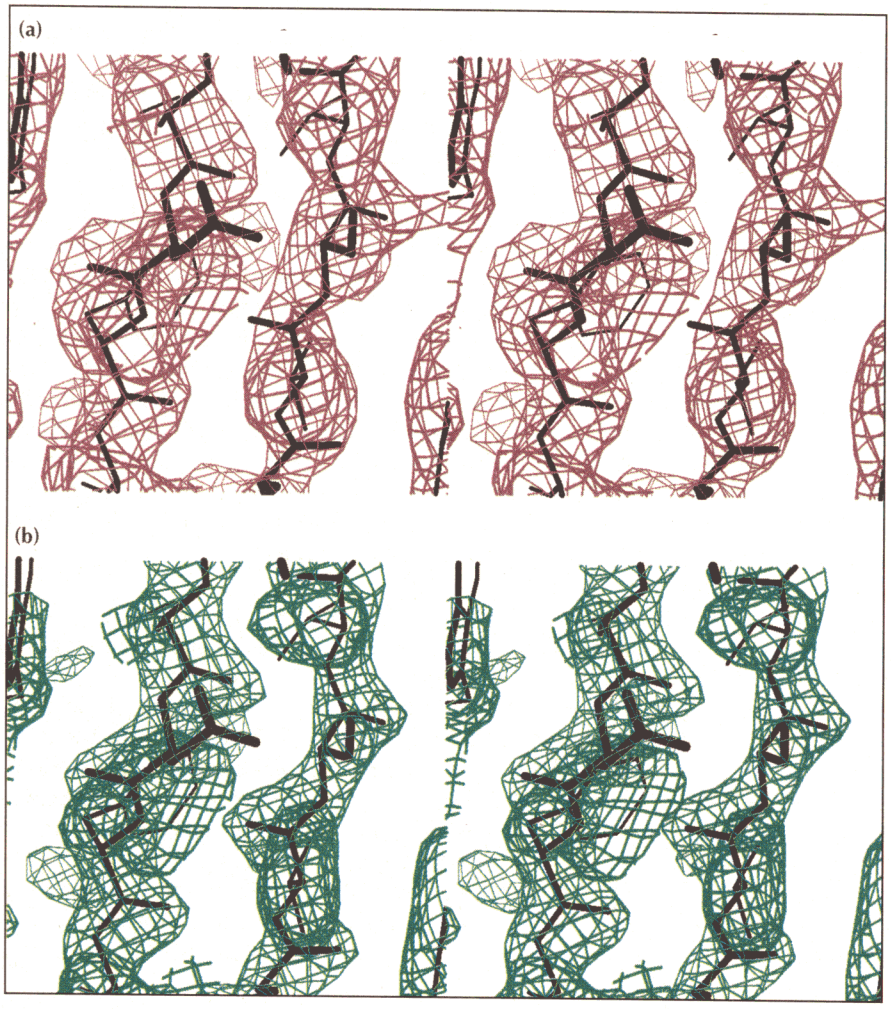


**Fig. 1.** Quality of the electron-density maps for ChiA. **(a)** Solvent-flattened MIR density in a well-defined region of the map in the amino-terminal domain, with the final model. **(b)** The same region with the final $3F_o$–$2F_c$ map. Drawn by O/OPLOT [39].

```
                L N A  r p g  s r h r l h
EMBL            CTGAATGCGC-GGCCTGGAA-GCCGACACCGCTTACACC

This work       CTGAATGCGCCGGCCTGGAAACCGGACACCGCCTACACC
      407       L N A P A W K P D T A Y T

                h g e r r e c a a g  Q G V
EMBL            ACGGTGAACGGCGTGAATGCGCTGCTGGC-CAGGGCGTC

This work       ACGGTGAACGGCGTCAATGCGCTGCTGGCGCAGGGCGTC
                 T V N G V N A L L A Q G V   432
```

**Fig. 2.** Comparison of two sequences for ChiA. The line labelled EMBL corresponds to the gene sequence from the EMBL databank (accesssion no. L01455 — it is the newest gene sequence from the two mentioned in the SWISSPROT protein entry). Above this line is the amino acid sequence translated from that sequence. Amino acids in lower-case letters indicate those that differ from the translation of the sequencing presented in this work. The sequence labelled 'This work' is the corrected sequence obtained in this study; the amino acids resulting from its translation appear below. After the first base insertion in the DNA sequence, the protein sequences are completely different, as a result of the frame shift. After the third insertion, the translation frame in the old sequence is 'corrected' and the old amino acid sequence reappears for both DNA sequences.

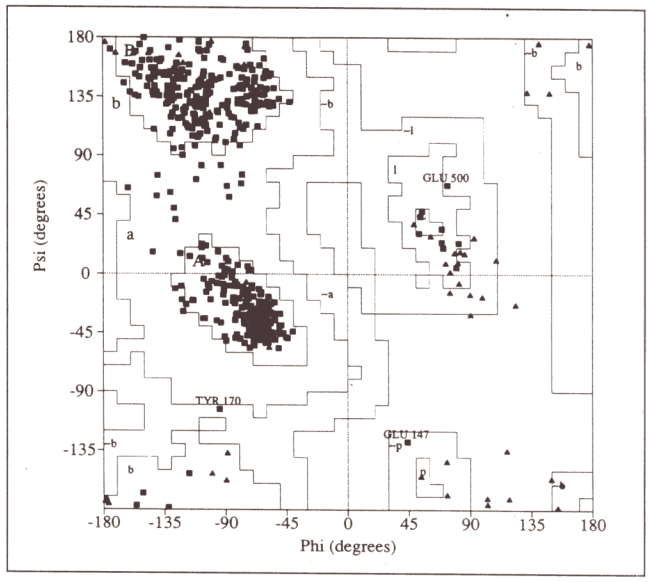


**Fig. 3.** Ramachandran plot for ChiA, created by PROCHECK [21]. Glycine and proline residues are represented as triangles and all other residues as squares. The different regions defined by borderlines are labelled as: A, B, and L, most favoured; a, b, l and p, allowed; and ~a, ~b, ~l and ~p, generously allowed; as defined in PROCHECK. The three residues in the generously allowed regions are labelled.

### Assessment of structural quality

The PROCHECK program [21] was used to assess the stereochemical quality of the model. PROCHECK is based on the analysis of 118 structures solved at resolution better than 2.0 Å and refined to an R-factor <20%. Most of the residues (92.1%) lie within the most favourable regions of the Ramachandran plot (Fig. 3). Assessment of the Ramachandran plot and peptide-bond planarity shows that these are better than typical values, whereas bad contacts, α-carbon tetrahedral distortion and hydrogen-bonding energies are well within the expected ranges. The overall G-factor, which indicates the stereochemical quality of the main chain, as defined in PROCHECK, is at the top of the expected values for this resolution. Fig. 4 shows four sets of structural data for ChiA plotted against residue number; the average B-factors of main-chain atoms; the average B-factors of side-chain atoms; the real-space correlation coefficients with the final $3F_o–2F_c$ map; and surface accessibilities. It can be clearly seen from the comparison of these four graphs that the most disordered areas are those exposed to the surrounding solvent. Some low real-space coefficients can be matched to disordered side chains, which are all located in loop regions.

### Overall structure

The structure consists of three domains. The amino-terminal domain (residues 24–137), which consists only of β-strands, connects through a hinge region (residues 138–158) to the main α/β-barrel domain (residues 159–442 and 517–563). The third domain, which has an α+β fold, is formed by an insertion in the barrel motif (residues 443–516) (see Fig. 5). The average B value for protein atoms is 24.1 $Å^2$.

#### *The amino-terminal domain*

The amino-terminal domain has a fold comprising only β-strands. This fold resembles the fibronectin III (FnIII) module domain [22] (Fig. 6a). The overall topology is similar, but a number of differences are evident. Firstly, strand A in FnIII-type domains, is long and makes a relatively large β-sheet together with strands B and E. In comparison, the same strand in ChiA is very short and only two of its residues are hydrogen bonded to strand B. Secondly, ChiA has a long insertion between strands A and B. This insertion forms first a small β-strand (A′) and then a small strand (A″)-helix–strand (B′) domain. Thirdly, strands C′ and C are linked by a cysteine bridge, which is absent in all of the similar folds. Finally, strand A′ of ChiA forms a small β-sheet together with the first part of strand F and the second part (G′) of strand G. A similar fold is also observed in the cytochrome *f* large domain [23] where A′ (which comes just after A) interacts with strand G. The result, in cytochrome *f*, is an additional strand to the existing β-sheet, formed by C, C′, F, G and A′. In that structure, a third, distinct β-sheet is formed, consisting of two antiparallel strands (the beginning of strand F and the whole of G′) and A′, which is parallel with G′. Strand F is not interrupted, but after a certain point it clearly belongs to the third β-sheet. This third sheet is completely absent from typical FnIII-type domains.

The sequence for the amino-terminal domain could be identified in two more chitinases from *Aeromonas caviae* (GENBANK, accession no. U09139) and *Alteromonas* species [24], using the BLASTP algorithm [25] as implemented in the NCBI BLAST network server using

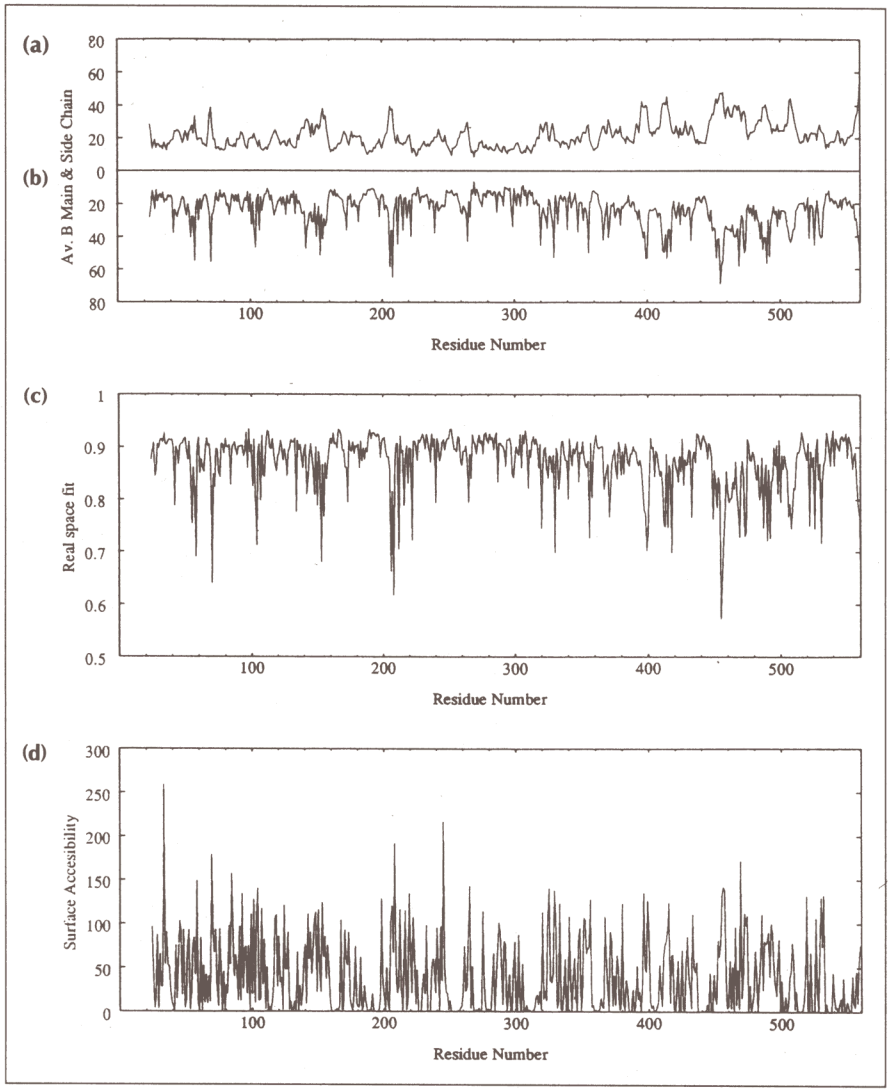


**Fig. 4.** Structural data for ChiA plotted as a function of residue number. Average B-factor for **(a)** main-chain atoms and **(b)** side-chain atoms. **(c)** Real-space correlation coefficient (as implemented in O [39]). **(d)** Surface accessibility.

NCSA Mosaic. The sequence for this domain has no similarities with the sequences of FnIII-type domains, or immunoglobulin constant and variable-type domains, although they have similar folds.

No biochemical data are available for the function of this domain. It has been shown that sequences for typical FnIII-type domains exist in three other bacterial chitinases [26]. FnIII domains can also be located in some cellulases. This fold, either formed by a typical FnIII-type domain or by the amino acid sequence presented here for ChiA, seems to have an important functional role; it does not seem to be an 'evolutionary relic' accidentally found in that particular chitinase. It is possible that this domain is involved in interaction with the filamentous chitin substrate.

*The hinge region*

The amino-terminal domain is connected to the rest of the structure via a 21-amino-acid hinge region (residues 138–158). This region was judged to have unusual characteristics, even before structure solution. These 21 amino acids comprise 4 prolines, 10 charged residues (including 5 lysines), 4 leucines, 1 tyrosine and 1 serine. This hinge region is probably mobile, allowing the amino-terminal domain to have different relative positions in solution. This is clearly reflected in the B-factors. The average B-factor for all main-chain atoms for the hinge region is 27.2 $Å^2$ whereas the average B-factor for the preceding amino-terminal domain is 18.6 $Å^2$ and for the rest of the α/β-barrel is 20.3 $Å^2$.

*The α/β-barrel*

The eight-stranded α/β-barrel is the catalytic domain of the enzyme, comprising residues 159–442 and 517–563 (Fig. 6b). It has a number of irregularities, compared with a typical α/β-barrel fold, such as are found in triose phosphate isomerase or pyruvate kinase (reviewed in [27]).

Firstly, strands B1 and B2 are connected not by an α-helix but by three short $3_{10}$-helices (G1-1, G1-2, and G1-3, each five residues long). Secondly, between strands B2 and B3 lie two α-helices (A′ and A2) as well as a $3_{10}$-helix (G2), which is seven residues long. Thirdly, between strands B3 and B4 a $3_{10}$-helix (G3)

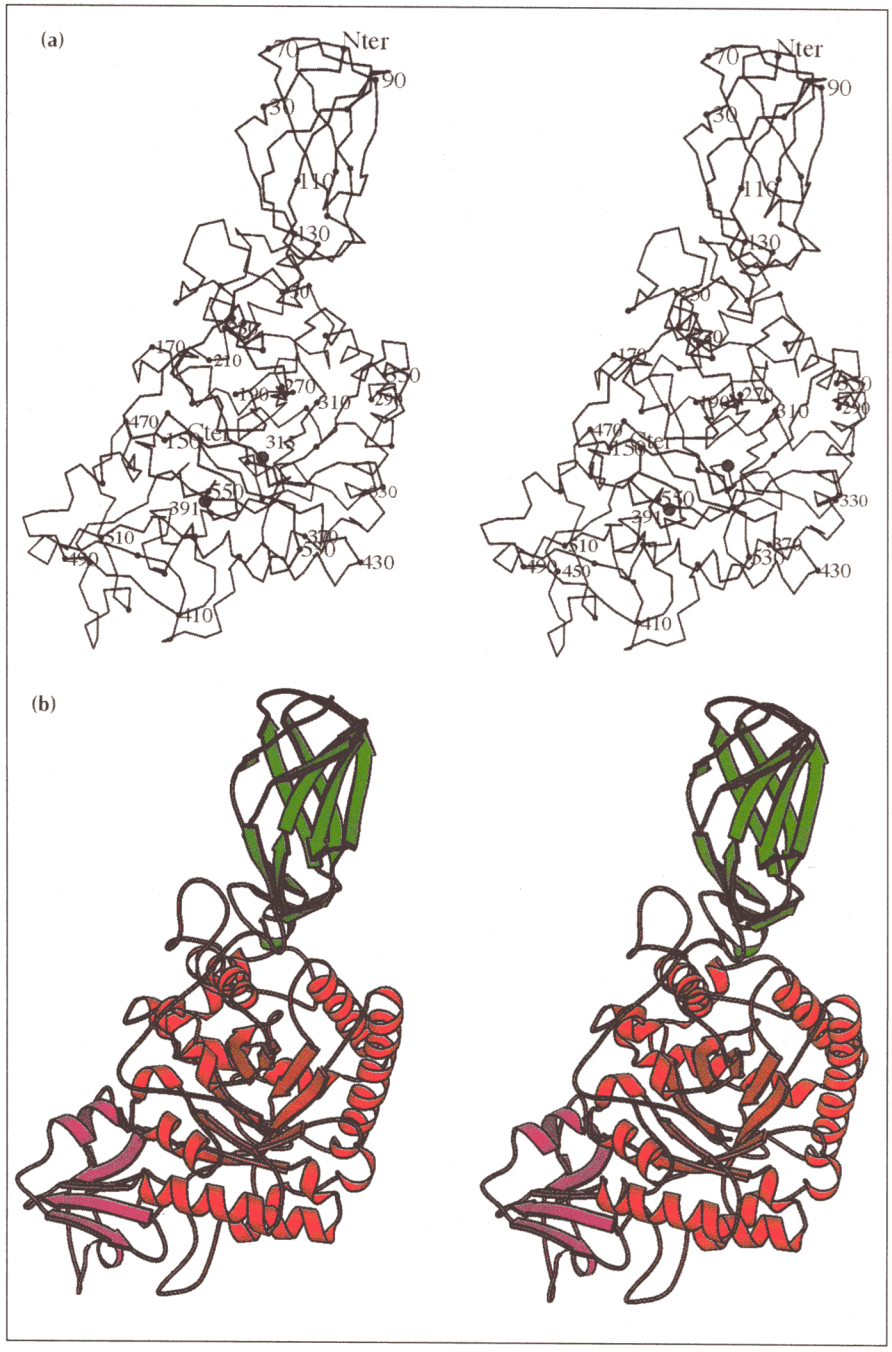


**Fig. 5. (a)** Stereo representation of the $C_\alpha$ trace of ChiA. Every 20th residue is numbered and every 10th residue has a black dot at the $C_\alpha$ position. The active-site residues are numbered and indicated by a larger dot. **(b)** Ribbon diagram of the structure in the same orientation as (a). The three domains are coloured differently; the amino-terminal domain in green, the α/β-barrel domain in red and the α+β domain in magenta. Both diagrams are drawn with MOLSCRIPT [44].

seven residues long is located in addition to the α-helix (A3). Fourthly, a relatively large insertion is found between B6 and A6, which forms two loops which 'protect' one side of the β-sheet of the α+β domain. Fifthly, A6 is followed by one short, four-residue $3_{10}$-helix (G6). Sixthly, the largest insertion is located between strands B7 and B8 and forms the third α+β domain. Finally, the last helix (A8) is preceded by a small five-residue $3_{10}$-helix (G8).

*The α+β domain*

The α+β domain is formed by an insertion between strand B7 and helix A7, (Fig. 6c). It comprises five β-strands (H, I, K, L and M), one of which (M) is interrupted. These make up an all antiparallel β-sheet. An α-helix protects the first hydrophobic surface of the sheet. The other surface of the sheet is protected by a $3_{10}$-helix together with some coil structure. This domain has a high average temperature factor (32.9 Å$^2$) compared with the rest of the structure. The average B-factor for the structure excluding this domain is 20.3 Å$^2$.

## Unusual geometries

*The $3_{10}$-helices*

$3_{10}$-Helices do occur in protein structure, although they are relatively rare compared with α-helices. For defining the $3_{10}$-helices and α-helices we used the criteria implemented in DSSP [28]. A search of the Protein Data Bank (PDB) [29] showed that out of all residues participating in a helical conformation, 12.6% belong to $3_{10}$-helices.

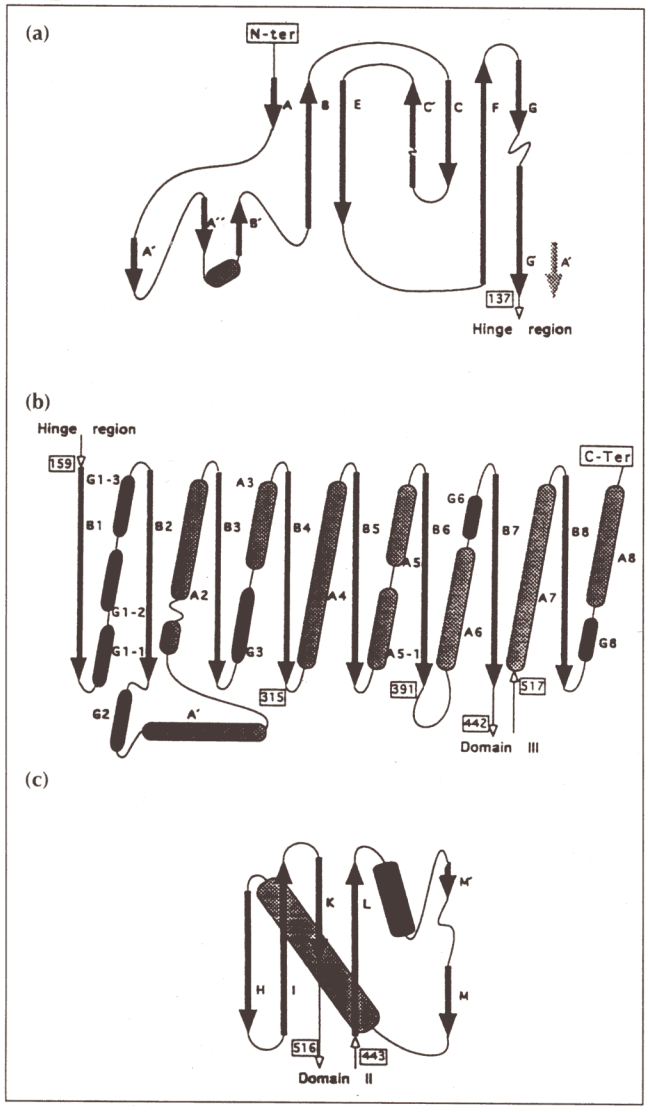


**Fig. 6.** Topology diagrams of the three ChiA domains: **(a)** the amino-terminal domain; **(b)** the α/β-barrel domain; and **(c)** the α+β domain. β-Strands are shown as arrows, α-helices as cylinders, and $3_{10}$-helices as darker and thinner cylinders. Starting and finishing residue numbers are indicated for each domain as are the positions of the active-site residues.

In ChiA 43 residues participate in $3_{10}$-helices, which represent 29% of the total residues found in helical conformation. Furthermore, the number of $3_{10}$-helices (8) is large compared with the number of α-helices (12). Most of the $3_{10}$-helices are close to the groove that facilitates the binding of the substrate.

### *A helix with an insertion*

The third helix (A2) between strands B2 and B3 starts with residues 230–232, the carbonyl oxygens of which are hydrogen bonded to the nitrogens of residues 234–236 respectively. Residues 233 and 234, however, have carbonyl oxygens hydrogen bonded to nitrogens of residues 252 and 253. The carbonyl oxygens of residues 250 to 259 are properly hydrogen bonded to the corresponding nitrogens of the $n+4$ residues. Residues 237–249 do not belong to the helix, but instead form a big loop. This results in the formation of an almost perfect helix without any break in the hydrogen-bonding pattern, as illustrated in Fig. 7.

### Cis *peptide bonds*

*Cis* peptide bonds have been previously reported for non-proline residues [30]. They are often associated with the active sites of enzymes in order to achieve essential constraints. Three *cis* peptide bonds were identified in ChiA. When these bonds were restrained during refinement to be in the usual *trans* conformation (ω=180°), the ω angle converged to values close to 0° (±20°). When these bonds were restrained to be in the *cis* conformation (ω=0°), ω converged at 0°. Omit maps were created for all three cases, and the electron density showed that these peptide bonds should be modelled in the *cis* conformation. The first *cis* peptide bond is between Gly190 and Phe191. The second is between Glu315 and Phe316. Three of these residues are clearly involved in the active site (discussed below). The third *cis* peptide bond occurs between residues Trp539 and Glu540. Trp539 is conserved in bacterial family 18 enzymes, and lies close to the active site, where it is probably involved in the binding of substrate. This is the only example of three *cis* peptide

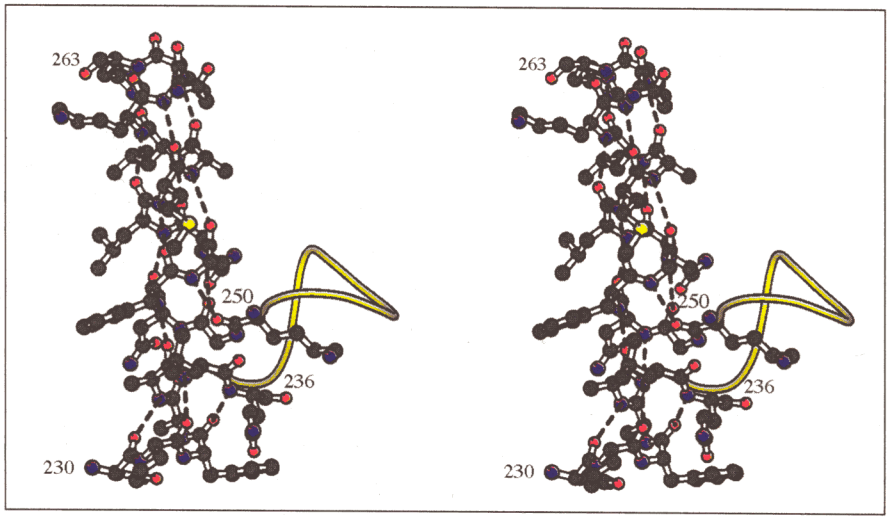


**Fig. 7.** Stereoview of a ball-and-stick model of the helix (A2) with the 14-amino-acid insertion. Hydrogen bonds are drawn as dotted lines. The insertion is drawn as a long tube. The hydrogen bonding pattern is consistent all over the helix and is only slightly distorted by the insertion. Start and end residues of the helix numbered. Figure drawn using MOLSCRIPT [44].

bonds in a protein where none of them are associated with a proline residue.

### The active site

The structure of ChiA in complex with *N,N′,N″,N‴*-tetra-acetylo-chitotetraose (tetra-NAG) showed one sugar ring bound to the active site (Fig. 8). The probable reason that only one ring is present is that chitinase cleaved the tetra-NAG molecule, suggesting that at pH close to optimal even oligosaccharides can be cleaved. An alternative explanation is that the remaining three rings are disordered and thus not visible in the electrondensity maps. No clear evidence for the exact positioning of the ring, by rotation along the axis vertical to the approximate ring plane, emerged upon inspecting the maps.

The substrate-binding site is formed by a long groove, located at the carboxy-terminal end of the β-strands of the α/β-barrel. In all known enzymes with α/β-barrel structure the active site is located at that end of the barrel (Fig. 9). Several aromatic residues are located along the groove as well as charged residues that can be presumed to form the subsites for the substrate binding, commonly named A–F, in analogy to lysozyme.

Several pieces of evidence point towards Glu315 and possibly Asp391 being located at the active site. Firstly,

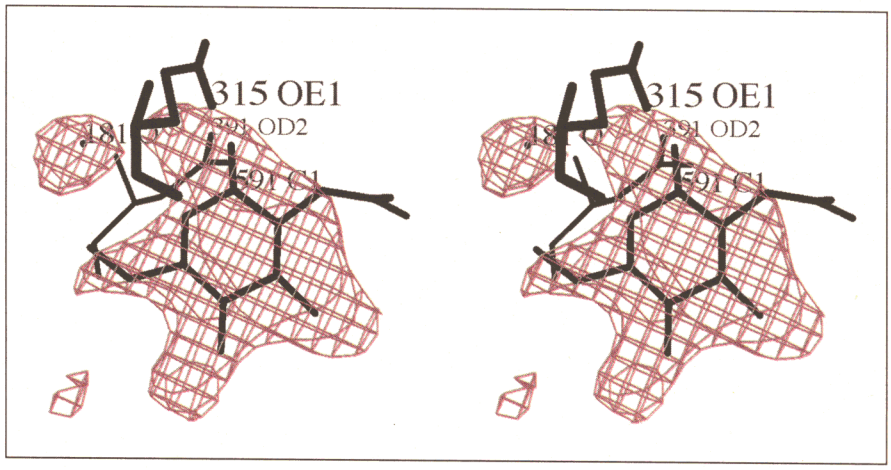


**Fig. 8.** Difference density ($F_o$–$F_c$, with phases from the native model) clearly showing the bound NAG. The active-site residues are shown. Density is contoured at the 4σ level. The orientation of the ring is not clear, and it is modelled in the most likely position according to chemical knowledge of the reaction mechanism, as described in the text. The role of water181 is unclear. It is presented here to show the interpretation of this density. Figure drawn using O/OPLOT [39].

**Fig. 9.** Ribbon diagram of ChiA in an orientation to show the three domains and the groove of the active site. The bound sugar is shown as a space-filling model (in magenta), and the two active site residues are drawn as a ball-and-stick model (in green). β-Strands are represented as purple arrows, α-helices as blue spirals, $3_{10}$-helices as orange spirals and coil structure and turns as yellow tubes. Most of the $3_{10}$-helices cluster around the groove of the binding site. Figure drawn by MOLSCRIPT [44].

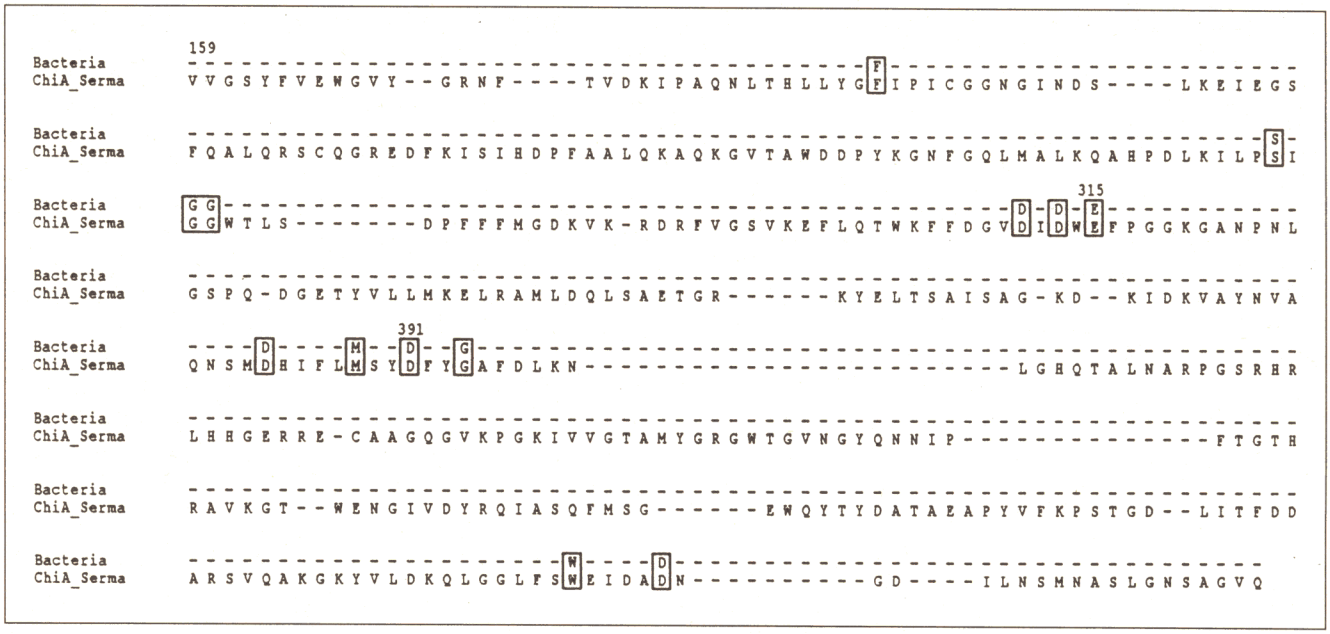


**Fig. 10.** Results of a multiple alignment of ChiA with 10 bacterial chitinase sequences from family 18 of the glycosyl hydrolases. The amino-terminal domain is omitted from this alignment. Only residues present in 9 or 10 out of the 10 sequences are shown for the bacterial consensus sequence (Bacteria); all other residues are indicated by dashes. Dashes in the *S. marcescens* ChiA sequence (ChiA-Serma) represent gaps introduced in the ChiA sequence during the alignment. The final alignment contains six conserved glutamate or aspartate residues. Asp383 and Asp544 are located far from the bound sugar and cannot be involved in catalysis. Residues Asp311 and Asp313 are not located in a favourable position in the groove. Hence, as described in the text, Glu315 and Asp391 have been identified as the catalytic residues and are numbered. The conserved sequence S-GG is not associated with the active site and its function is not yet known. Alignment was performed using CLUSTALW [45], consensus calculated using GCG/PILEUP and alignment drawn by EMBL GCG/PRETTYPLOT.

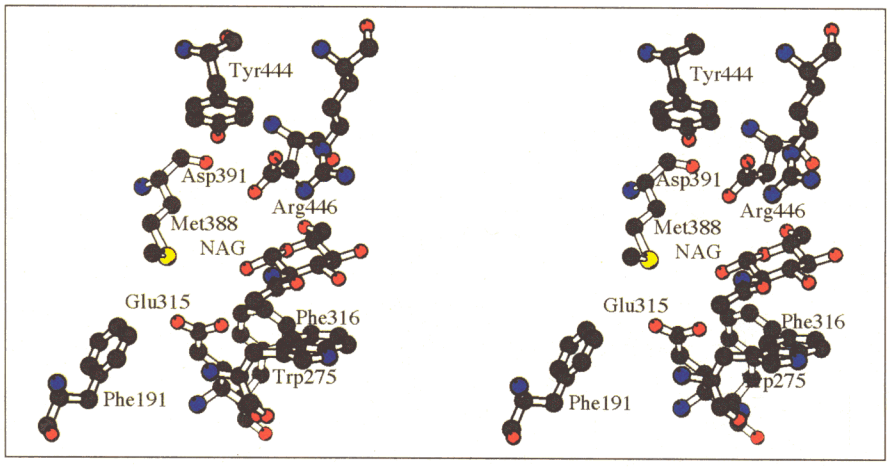


**Fig. 11.** Stereo representation of the active site of ChiA. Carbons are drawn in black, oxygens in red and nitrogens in blue. Details for the proposed mechanism are given in the text.

site-directed mutagenesis in the *Bacillus circulans* chitinase [31] showed that the Glu204→Gln mutation (Glu204 of *Bacillus* chitinase aligns with Glu315 of ChiA) decreased activity almost to zero. No mutation was made in the residue corresponding to Asp391. Mutation of Asp200 (which corresponds to Asp311 in ChiA) to asparagine decreased $k_{cat}$ 350-fold and could not be shown to be directly associated with the reaction mechanism. Secondly, Glu315 is completely conserved in family 18 chitinases. Asp391 is conserved in nine out of ten bacterial chitinases of family 18 (Fig. 10). Thirdly, the sugar ring was modelled with the O1 atom orientated towards Glu315, which is the first active-site residue, as shown by the biochemical data. In that orientation, one of the Asp391 carboxylate oxygens is close to the C1 atom of the ring (Fig. 8). Asp391 is the only charged residue, other than Glu315, in the area where the sugar ring is bound. The closest carboxylate oxygens of Asp391 and Glu315 are 6.9 Å apart and in two almost symmetrical positions with respect to the plane of the bound sugar ring. This resembles the case of the active site residues of lysozyme [32] where the distance is 6.2 Å and the geometry is comparable.

It cannot be proved that the bound sugar is modelled in the correct orientation. The alternative would be that this sugar is the product of the reaction and bound in the 'product' subsite E instead of the 'substrate' subsite D. In

that case, the sugar ring should be modelled with C4 in the current position of C1. The latter is unlikely, firstly because the product does not usually stay bound after the completion of the reaction, and secondly, poor density for a second sugar ring (not sufficient to model it) indicates that a second disaccharide is bound, or at least partly bound to the C4 of the current model. If the first sugar is modelled as the product of the reaction, then we would expect a disaccharide to be bound in the product site. That would leave all the substrate sites free, which is unlikely. Unfortunately, the electron density is not clear enough to demonstrate the exact position of the ring and show clearly the mode of binding or explain the enzyme specificity.

Provided that the sugar is modelled correctly and reasonably accurately, the following assumptions can be made about the catalytic mechanism. Glu315 is in a hydrophobic environment, surrounded by Phe191, Phe316, Met388 and Trp275 (Fig. 11). At acidic pH it would thus be expected to be protonated and should be the proton donor in the general acid-base catalytic mechanism. The oxygen of Glu315 is only 2.9 Å away from O1. This distance is almost identical to that separating the corresponding atoms in the lysozyme complex (the Glu35–O1 distance is 2.8 Å). Asp391 is in a more hydrophilic environment, surrounded by Tyr444 and Arg446 and is likely to have a negative charge at that pH, which would allow it to stabilize the carboxy-anion intermediate. The oxygen of Asp391 is placed 3.2 Å from the C1 atom. This distance is the same as that which separates Asp52 from C1 in lysozyme (3.2 Å).

According to the stereochemistry of the active site, the cleavage of chitin by ChiA is likely to proceed by general acid-base catalysis, in a similar manner as in lysozyme and most glycosyl hydrolases. The above observations are consistent with the evidence that for *B. circulans* the cleavage of the glycosidic bond by chitinase proceeds with retention of configuration of the anomeric conformation of the C1 atom [33], as in lysozyme. The exact mode of binding of the sugars in the groove and the reasons for the specificity for chitin remain to be answered.

## Biological implications

**Chitinases (EC 3.2.1.14) hydrolyze the chitin homopolymer, one of the most abundant natural polymers, into smaller oligomeric saccharides, comprising *N*-acetyl-D-glucosamine units. This compound is the main structural component of the cell wall of most fungi and of the exoskeleton of arthropods. Chitinases are widespread in nature and have been found in bacteria, fungi, plants, invertebrates and vertebrates.**

**Bacterial chitinases are important biological control agents for countering fungal infections of plants some of which are of major economic interest. Biotechnological progress on this front requires structural data, but although crystal structures have been reported for two prototypic plant enzymes, no structure has been reported for bacterial or fungal enzymes.**

**We have begun to address this shortfall by determining the crystal structure of a bacterial chitinase, chitinase A (ChiA), from the soil bacterium *Serratia marcescens*. It comprises three domains, of which the largest is the catalytic α/β-barrel domain. The active site was identified by solving the structure of the enzyme in the presence of an oligomer of its natural substrate. The poor quality of the complex prevents both clear insights into the mode of substrate binding and identification of the structural features that determine the specificity for the chitin polysaccharide. The catalytic residues include a glutamate and possibly an aspartate consitent with a catalytic mechanism similar to that of lysozyme, i.e. general acid-base catalysis.**

**The amino-terminal domain has a fold similar to that of the animal protein fibronectin type III module domains. Its function is as yet unknown but might well facilitate the binding of the enzyme to the filamentous chitin substrate.**

**This structure takes us one step nearer a complete understanding of the specificity and selectivity of the glycosyl hydrolases for slightly different natural carbohydrate polymers. The present results provide the basis for detailed exploration of the catalytic mechanism.**

## Materials and methods

### *DNA sequencing*

A 400 bp long fragment, containing the problematic sequence, was amplified by PCR and cloned in the pcrII™ vector. Positive clones were identified by restriction analysis, using two external restriction sites. Double-stranded DNA sequencing was performed by the DNA sequencing service at European Molecular Biology Laboratory (EMBL), Heidelberg. Sequencing of both strands had no compressions and was clearly readable.

### *Crystallization*

Crystallization of ChiA has been described in [34]. New conditions, resulting in crystals diffracting to significantly higher resolution, are presented here. The vapour diffusion method was used. The protein was stored in 20 mM Tris-HCl, at pH 8.0. Hanging drops of protein solution were equilibrated against 1 ml of reservoir buffer. The drop contained 40–50 μg of protein and 50% v/v reservoir buffer. The reservoir buffer consisted of 20–22% PEG-4000, 0.15–0.2 M ammonium sulphate, 0.5–1% v/v isopropanol and 0.1 M acetate buffer, pH 5.2. The best crystals were obtained after 1–2 weeks. They were flat plates with dimensions 0.5 mm×0.5 mm×0.05 mm. All crystallization trials were carried out at 17°C. Crystals of the complex with the substrate were prepared by soaking for 24 h in a solution of 50 mM tetra-NAG in 30% PEG-4000 and ammonium sulphate, isopropanol and buffer as indicated for the reservoir buffer.

### *Data collection*

All data were collected at EMBL beamline X11 located at HASYLAB, DESY Hamburg. Native data were collected to a maximum resolution of 2.3 Å. Data for the complex with tetra-NAG were collected to 2.5 Å resolution. In all cases, cooling to 4°C was used. Both native and complex data were collected on a single crystal. No significant radiation damage was observed. Special care was taken during data collection for the second Pb derivative, in order to collect good anomalous data. The crystal was aligned so as to collect both Friedel pairs in one frame. The wavelength was 0.92 Å, below the absorption edge for Pb to maximize the anomalous scattering.

All data were indexed and scaled using the DENZO [35] and SCALEPACK [36] programs respectively. For the derivatives, Friedel opposites were kept unmerged and used for the scaling. Statistics on the native and complex data are shown in Table 1. Intensities were truncated to amplitudes by the TRUNCATE program [37].

For the anomalous data the TRUNCATE program was not used. Instead, all reflections for which both Friedel opposites were measured and had non-negative intensities were kept. The amplitudes were computed as the square root of the intensities together with their σ values. Friedel opposites were thus kept separately and used for phasing.

**Table 1.** Statistics for the crystallographic data collection for the native enzyme and its complex with tetra-NAG.

| Chitinase A crystal statistics | | Native and complex data | | |
|---|---|---|---|---|
| | | | **Native** | **Complex** |
| MW (kDa) | 62 | | | |
| a (Å) | 203.1 | Resolution (Å) | 2.3 | 2.5 |
| b (Å) | 133.9 | $R_{merge}$ (%) | 7.7 | 10.1 |
| c (Å) | 59.9 | $R_{merge}$ (last shell) (%) | 24.6 | 28.1 |
| $\alpha = \beta = \gamma$(°) | 90 | I/σI | 12.5 | 14.0 |
| Space group | $C222_1$ | I/σI (last shell) | 5.1 | 3.4 |
| $V_m$ | 3.2 | Unique reflections | 38 115 | 28 296 |
| Solvent | | Redundancy | 5.2 | 7.1 |
| content (%) | 54 | Completeness (%) | 97.2 | 98.4 |

$R_{merge} = \Sigma_h[(\Sigma_I|I_I - \langle I \rangle|)/\Sigma_I|I|]$. Redundancy = measured reflections/unique reflections.

### *MIR phasing and phase improvement*

All scaling of native and derivative data, Patterson map calculation, heavy-atom refinement, solvent flattening and map calculations were performed using the programs included in the PHASES package [38]. The difference Patterson map for the first Pb derivative clearly showed one strong site and was straightforward to solve. For all other derivatives, difference Fourier techniques using the phases obtained from the first derivative were used. All the sites that were found and used for the final phasing were clearly visible in the difference Patterson maps. For the conventional heavy-atom parameters refinement only phases with figure of merit (FOM) better than 0.6 were used. During refinement of the parameters of one derivative, the contribution of this derivative to the phasing was used. The reason for this was only one derivative was available with a resolution better than 4.5 Å. The final FOM was 0.66. Statistics for the quality of the derivative data and the phasing are shown in Table 2.

The resulting MIR map was solvent flattened with the procedure implemented in the PHASES package. Solvent flattening improved the density dramatically. No other phase extension or modification was performed. A skeleton model was calculated with the latter package.

The high quality of the second Pb derivative data set provided excellent anomalous phasing information, as is clearly shown in the phasing statistics. These are the only derivative data contributing to better than 3.2 Å. Together with the first Pb data set they are the only ones used for phasing to resolution better than 4.5 Å.

### *Model building and refinement*

Model building was done by the O program [39]. After editing the bones, $C_\alpha$ positions were assigned by the BATON option. A polyalanine model was built using the Lego commands, performing a database search to obtain the best-fitting pentapeptide for the given $C_\alpha$ conformation. The sequence was then assigned using the 'slider guess' and 'slider combine' options, which proved to be extremely useful. Side chains were put in the most common rotamer position and then manually inspected and fitted into density using only the most common rotamers. In this way 481 residues were built into the initial solvent-flattened MIR map.

Refinement of the model was performed with the standard X-PLOR simulated annealing protocol [40]. The crystallographic

**Table 2.** Statistics for the data quality for the derivatives and phasing.

| Derivative data | | | | | Phasing statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Derivative | Resolution (Å) | $R_{merge}$(%) | Completeness (%) | $R_{iso}$[a] (%) | Derivative | No. of sites | Phasing power[b] | $R_{cullis}$[c] (%) | $R_{Kraut}$[d] (%) |
| Pb1 | 3.2 | 6.5 | 59.6 | 16.9 | Pb1 | 1 | 2.08 | 68 | 12.9 |
| Pb2 | 2.4 | 3.4 | 96.2 | 25.7 | Pb2/Pb2 (ano) | 1/1 | 1.99/2.28 | 64/– | 8.6/13.4 |
| Pt1 | 4.5 | 5.9 | 99.0 | 27.9 | Pt1/Pt1 (ano) | 2/2 | 1.38/1.51 | 64/– | 16.3/19.4 |
| Pt2 | 4.5 | 8.5 | 96.4 | 30.9 | Pt2 | 2 | 1.31 | 66 | 17.8 |

Overall figure of merit (10–2.4 Å) = 0.66. Pb1 and Pb2: trimethyl-lead-acetate derivatives; for Pb1, crystals were soaked for 2 days in 5 mM heavy-atom solution; and for Pb2, crystals were soaked for 1 week; the binding site lies between Arg343 and Asp383. Pt1, platinum tetrachloride derivative; crystals were soaked in 2 mM heavy-atom solution for 6 h. Pt2: platinum (II)-(2,2′-6′,2″-tetrapyridin) chloride derivative; crystals were soaked in 2 mM heavy-atom solution for 12 h. The two sites, common for both, are located with one at His229 and the second between Asp391 and Arg446. [a]$R_{iso} = (\Sigma_h|FP_h - FPH_h|)/(\Sigma_h FP_h)$. [b]Phasing power = rms (FH/E), where $E = ||FPH_{obs}| - |FPH_{calc}||$. [c]$R_{cullis} = (\Sigma_h|FPH_{obs} \pm FP_{obs} - FH_{calc}|)/(\Sigma_h|FPH_{obs} \pm FP_{obs}|)$ for centric reflections. [d]$R_{Kraut} = (\Sigma_h|FPH_{obs} - FPH_{calc}|)/(\Sigma_h|FPH_{obs}|)$ for acentric reflections.

R-factor converged from 52.3% to 32.1% and $R_{free}$ [41] to 36.2%. Phase combination of model and solvent-flattened phases was done with the PHASES package. The resulting map revealed all the missing residues except the two carboxy-terminal ones. All residues were built into this map except for the 20 residues for which the density was clear for both main-chain and side-chain atoms, but it did not match the sequence. The density suggested that there should be 21 residues in this area. The density in this region did not match any other part of the sequence, as was shown with the 'slider guess' option. These residues were included as alanines at this stage. After refinement of this model the R-factor was 26.5% and $R_{free}$ 31.6%. A $3F_o$–$2F_c$ map using calculated phases showed again that the 21 residues that were built as alanines had different side chains from the ones suggested by the available DNA sequence.

At that point, these residues were incorporated into the model with the side chains deduced from the density. The R-factor converged to 23.1% with this model and $R_{free}$ to 28.8%, without including any waters.

Resequencing of the DNA of the ambiguous region showed that the map interpretation was indeed correct and the previous sequence was in error.

Water molecules were selected with the ARP package [42]. They were inserted where the $3F_o$–$2F_c$ difference Fourier map showed significant density, which was in all cases above 3.5σ. After refinement, the program removed all waters in the $F_o$–$F_c$ map which were in less than 1σ density. Thirty cycles of this procedure were used.

The final model was subjected to 10 cycles of Konnert-Hendrickson [43] least-squares refinement as implemented in the PROLSQ program from the CCP4 suite [SERC (UK) Collaborative Computer Project 4, Daresbury Laboratory, UK, 1979]. All waters with B-factors >80 Å$^2$ were removed from the model. The final model contained 332 water molecules. The crystallographic R-factor for the final model is 16.2%. Statistics on the refinement are shown in Table 3.

### *The complex structure*

Native crystals were soaked for 24 h with tetra-NAG. Data were collected to a resolution of 2.5 Å. The cell parameters were not significantly different from the native crystal. An $F_o$–$F_c$ Fourier map was calculated using $F_c$ and phases obtained from the native model after removing all water molecules. The difference density contoured at 4σ clearly showed the position of one sugar ring. The exact orientation of the ring was not clear from that density. When the density was contoured at 2σ another sugar ring adjacent to the previous one could be seen, but because of low occupancy, high disorder or both, it was impossible to model it. The first sugar ring was modelled to the density as shown and discussed in the previous sections. The face of the sugar ring is packed against the ring of Trp275. Refinement of this model with the PROLSQ program resulted in an R-factor of 21.0% including no water molecules. The inclusion of 220 water molecules, using the same procedure as described for the native structure resulted in a decrease of the R-factor to 16.0%. More statistics are shown in Table 3. The B-factors of the sugar atoms were in the range 55–65 Å$^2$, indicating partial occupancy or disorder, both possibly attributable to the method used to obtain the complex.

**Table 3.** Statistics for the refinement of the native and complex structures.

| | Native | Complex |
|---|---|---|
| Resolution (Å) | 2.3 | 2.5 |
| No. of residues | 538 | 538 |
| Missing residues | 2 | 2 |
| No. of protein atoms | 4122 | 4137 |
| No. of solvent atoms | 332 | 220 |
| No. of reflections | 38 115 | 28 296 |
| R (all reflections) (%) | 16.2 | 16.0 |
| R ($F > 3\sigma F$) (%) | 15.3 | 15.1 |
| R (excluding solvent) (%) | 22.8 | 21.0 |
| Average B-factor (protein atoms) (Å$^2$) | 24.1 | 18.5 |
| Average B-factor (all atoms) (Å$^2$) | 25.2 | 19.4 |
| Rms bond distances (Å) | 0.015 | 0.018 |
| Rms bond angles (Å) | 0.037 | 0.038 |

*Acknowledgements*: We are grateful to: Morten Kjeldgaard for help with map interpretation with O and the time he gave us on his Graphics workstations in Aarhus University; Bill Furey for help with his PHASES package; Victor Lamzin for help with his ARP package and other stages of refinement; Gerard Kleywegt for turbo-e-mail help with O; Ann-Marie Voie and the DNA sequencing service at EMBL, Heidelberg; Christos Ouzounis for help in sequence analysis and alignments; Gideon Davies for all the helpful discussions; Howard Terry for help with computer software and reading this manuscript; and Wojciech Rypniewski for useful suggestions.

## References

1. Jeuniaux, C. (1993). Chitinolytic systems in the digestive tract of vertebrates: a review. In *Chitin Enzymology*. (Muzzarelli, R.A.A., ed), pp. 233–244, European Chitin Society, Italy.
2. Spindler-Barth, M. (1993). Hormonal regulation of chitin metabolism in insect cell lines. In *Chitin Enzymology*. (Muzzarelli, R.A.A., ed), pp. 75–82, European Chitin Society, Italy.
3. Lamb, C.J., Lawton, M.A., Dron, M. & Dixon, R.A. (1989). Signals and transduction mechanisms for activation of plant defenses against microbial attack. *Cell* **56**, 215–224.
4. Legrand, M., Kauffmann, S., Geoffroy,P. & Fritig, B. (1987). Biological function of pathogenesis-related proteins: four tobacco pathogenesis-related proteins are chitinases. *Proc. Natl. Acad. Sci. USA* **84**, 6750–6754.
5. Collinge, D.B., *et al.*, & Vad, K. (1993). Plant chitinases. *Plant J.* **3**, 31–40.
6. Adams, D.J., *et al.*, & Dada, J. (1993). Regulation of chitin synthase and chitinase in fungi. In *Chitin Enzymology*. (Muzzarelli, R.A.A., ed), pp. 15–25, European Chitin Society, Italy.
7. Cabib, E., Silverman, S.J. & Shaw, J.A. (1992). Chitinase and chitin synthase: counter-balancing activities in cell separation of *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* **138**, 97–102.
8. Henrissat, B. (1991). Classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.
9. Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788.
10. Perrakis, A., Wilson, K.S., Chet, I., Oppenheim, A.B. & Vorgias, C.E. (1993). Phylogenetic relationships of chitinases. In *Chitin Enzymology*. (Muzzarelli, R.A.A., ed), pp. 417–422, European Chitin Society, Italy.
11. Hart, P.J., Monzingo, A.F., Ready, M.P., Ernst, S.R. & Robertus J.D. (1993). Crystal structure of an endochitinase from *Hordeum vulgare* 1. seeds. *J. Mol. Biol.* **229**, 189–193.
12. Terwisscha van Scheltinga, A.C., Kalk, K.H., Beintema, J.J. & Dijkstra, B.W. (1994). Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. *Structure* **2**, 1181–1189.
13. Roberts, R.L. & Cabib, E. (1982). *Serratia marcescens* C chitinase: one-step purification and use for the determination of chitin. *Analyt. Biochem.* **127**, 402–412.

14. Jones, J.D.G., Grady, K.L., Suslow, T.V. & Bedbrook, J.R. (1986). Isolation and characterization of genes encoding two chitinase enzymes from *Serratia marcescens*. *EMBO J.* **5**, 467–473.
15. Fuchs, R.L., McPherson, S.A. & Drahos, D.J. (1986). Cloning of a *Serratia marcescens* gene encoding chitinase. *Appl. Environ. Microbiol.* **51**, 504–509.
16. Oppenheim, B.A., Hirsch, Y., Koby, S. & Chet, I. (1990). Protein secretion in biotechnology. In *Biologicals From Recombinant Microorganisms and Animal Cells*. (White, M.D., Reuveny, S. and Shafferman, A., eds), Vch Weinheim, Germany.
17. Vorgias, C.E., Tews, I., Perrakis, A., Wilson, K.S. & Oppenheim, A.B. (1993). Purification and characterization of the recombinant chitin degrading enzymes, C chitinase A and C hitobiase from *Serratia marcescens*. In *Chitin Enzymology*. (Muzzarelli, R.A.A., ed), pp. 417–422, European Chitin Society, Italy.
18. Shapira, R., Ordentlich, A., Chet, I. & Oppenheim, A.B. (1989). Control of plant diseases by chitinase expressed from cloned DNA in *Escherichia coli*. *Phytopathology* **79**, 1246–1249.
19. Jach, G., *et al.*, & Logemann, J. (1992). Expression of a bacterial chitinase leads to improved resistance of transgenic tobacco plants against fungal infection. *Biopractice* **1**, 33–40.
20. Bairoch, A. & Boeckmann, B. (1993). The SWISSPROT protein sequence data bank, recent developments. *Nucleic Acids Res.* **21**, 3093–3096.
21. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
22. Leahy, D.J., Hendrickson, W.A., Aukhil, I. & Erickson, H.P. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science*. **258**, 987–991.
23. Martinez, S.E., Huang, D., Szczepaniak, A., Cramer, W.A. & Smith, J.L. (1994). Crystal structure of chloroplast cytochrome *f* reveals a novel cytochrome fold and unexpected heme ligation. *Structure* **2**, 95–105.
24. Tsujibo, H., *et al.*, & Inamori, Y. (1993). Cloning, sequence, and expression of a chitinase gene from a marine bacterium, *Altermonas sp.* strain O-7. *J. Bacteriol.* **175**, 176–181.
25. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipmann P.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
26. Bork, P. & Doolittle, R.F. (1992). Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA* **89**, 8990–8994.
27. Muirhead, H. (1983). Triose phosphate isomerase, pyruvate kinase and other αβ-barrel enzymes. *Trends Biochem. Sci.* **8**, 326–330.
28. Kabsch, W & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
29. Bernstein, F.C. & Tasumi, M. (1978). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
30. Herzberg, O. & Moult, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins* **11**, 223–229.
31. Watanabe, T., *et al.*, & Tanaka, H. (1993). Identification of glutamic acid 204 and aspartic as essential residues for chitinase activity. *J. Biol. Chem.* **268**, 18567–18572.
32. Strydanka, N.C.J. & James, M.N.G. (1991). Lysozyme revisited: crystallographic evidence for distortion of an *N*-acetylmuramic acid residue bound in site D. *J. Mol. Biol.* **220**, 401–424.
33. Armand, S., Tomita, H., Heyraud, A., Gey, C., Watanabe, T. & Henrissat, B. (1994). Stereochemical course of the hydrolysis reaction catalyzed by chitinases A1 and D from *Bacillus circulans* WL-12. *FEBS Lett.* **343**, 177–180.
34. Vorgias, C.E., Kingswell, A.J., Dauter, Z. & Oppenheim, A.B. (1992). Crystallization of recombinant chitinase from the cloned *chiA* gene of *Serratia marcescens*. *J. Mol. Biol.* **226**, 897–898.
35. Otwinowski, Z. (1993). *DENZO: an oscillation data processing program for macromolecular crystallography*. Yale University, New Haven, CT.
36. Otwinowski, Z. (1993). *SCALEPACK: Software for the scaling together of integrated intensities measured on a number of separate diffraction images*. Yale University, New Haven, CT.
37. French, S. & Wilson, K.S. (1978). On the treatment of negative intensity observations. *Acta. Crystallogr. A* **34**, 517–525.
38. Furey, W. & Swaminathan, S. (1990). PHASES — a program package for the processing and analysis of diffraction data from macromolecules. *American Crystallographic Association Meeting Abstracts*, **18**, 73.
39. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
40. Brünger, A.T. (1988). Crystallographic refinement by simulated annealing: application to a 2.8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.* **203**, 803–816.
41. Brünger, A.T. (1993). Assesment of phase accuracy by cross validation: the free R value. Methods and application. *Acta Crystallogr. D* **49**, 24–36.
42. Lamzin, V.S. & Wilson,K.S. (1993). Automated refinement of protein models. *Acta Crystallogr. D* **49**, 129–147.
43. Konnert, J.H. & Hendrickson, W.A. (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Crystallogr. A* **36**, 344–350.
44. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plot of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
45. Higgins, D.G., Bleasby, A.J. & Fuchs, R. (1992). CLUSTALV: improved software for multiple sequence alignment. *CABIOS* **8**, 189–191.


Received: **10 Aug 1994**; revisions requested: **29 Sep 1994**; revisions received: **10 Oct 1994**. Accepted: **17 Oct 1994**.